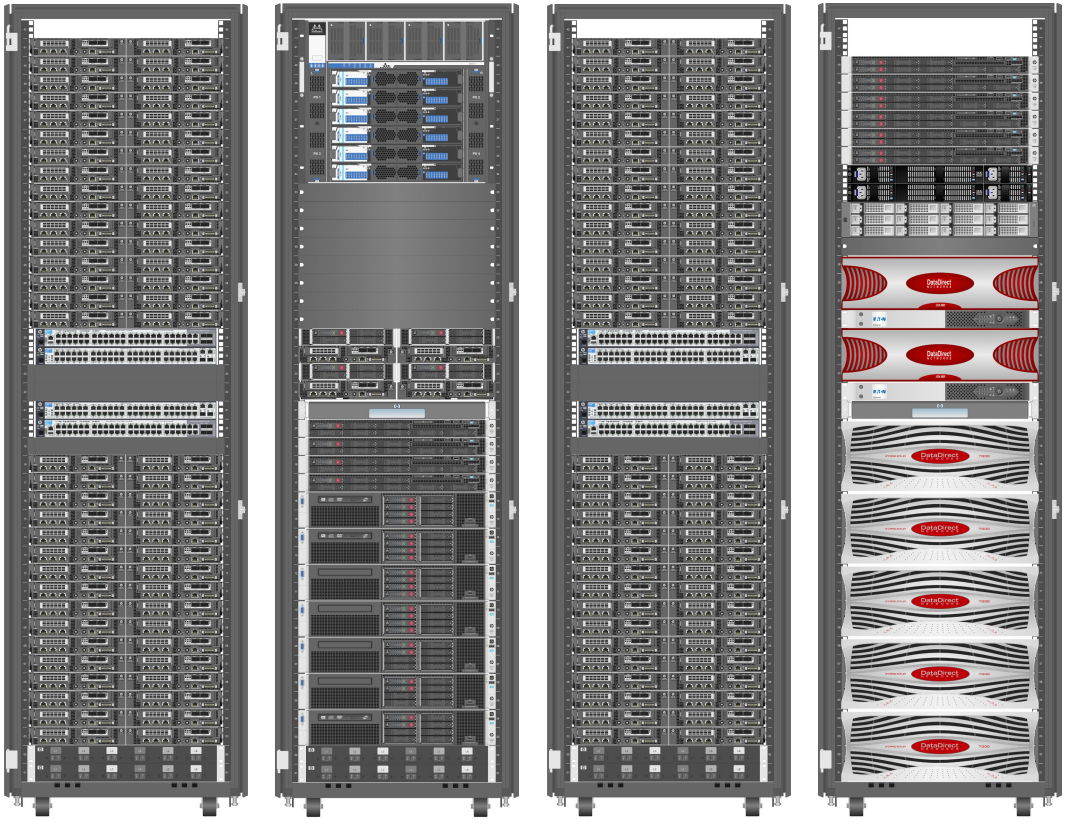
NLHPC
National Laboratory for High Performance Computing Chile

Guía básica para el uso de Leftraru

Arquitectura Leftraru



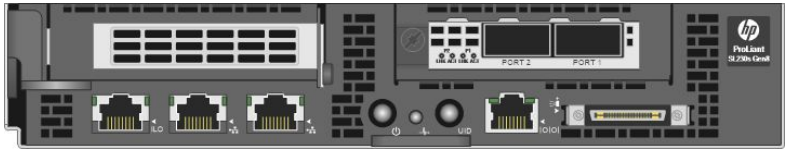
- Infiniband FDR 56Gbps
- Red iLO para adm. hw.
- Red servicio 1 Gbps
- +200 Tb almacenamiento DDN Lustre
- 128 nodos *slims*
- 4 nodos *fats*
- 12 Xeon Phi
- Racks enfriados por agua
- Enfriamiento in-row respaldo
- UPS 120 KVA autonomía: 30 mins.

NLHPC
National Laboratory for High Performance Computing
Chile

Nodos de cómputo Leftraru

2 tipos de nodo de cómputo:

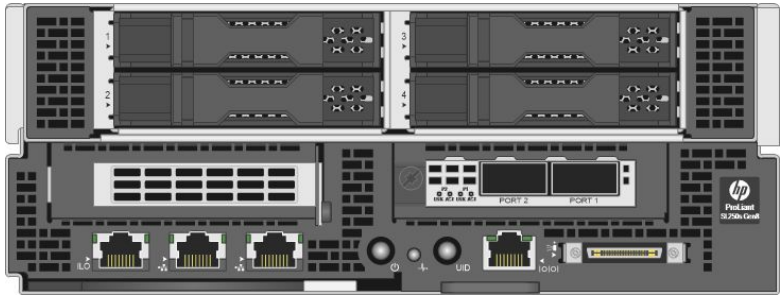
Slim



Nodo Slim

- 2 CPUS Xeon E5-2660 v2, 10 cores c.u.
- 48 GB RAM DIMM DDR3
- HD Interno 300 GB
- InfiniBand FDR 56 Gbps

Fat

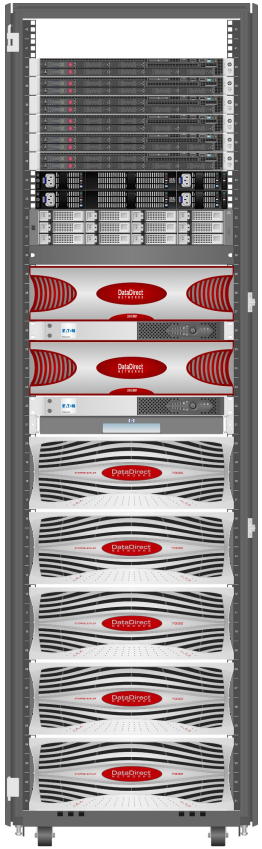


Nodo Fat

- 2 CPUS Xeon E5-2660 v2, 10 cores c.u.
- 64 GB RAM DIMM DDR3
- HD Interno 300 GB
- InfiniBand FDR 56 Gbps
- 3 coprocesadores Xeon Phi
  - 240 cores c.u.
  - 8 GB RAM c.u.

NLHPC
National Laboratory for High Performance Computing Chile

Almacenamiento Leftraru



DDN EXAScaler

- Almacenamiento paralelo de clase mundial
- Alto rendimiento en operaciones IO
- Tolerante a fallas (alta disponibilidad)
- Interconexión infiniband
- Capacidades Big Data

Características en Leftraru

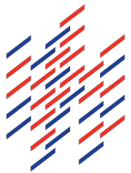
- Sistema archivos EXAScaler (Lustre)
- +200 TB de almacenamiento
- Almacenamiento metadata separado
- 2 controladoras SFA en H.A.
- 4 nodos OSS conectados a Infiniband
- 2 nodos MDS en H.A.

NLHPC
National Laboratory for High Performance Computing Chile

Introducción a SLURM

Simple Linux Utility for Resource Management

- Administra clusters y ejecución de tareas
- Open source
- Utilizado en el 60% de los supercomputadores del top500
- Versión actual instalada en Leftraru: 16.05.4



NLHPC
National Laboratory
for High Performance
Computing
Chile

Características de SLURM

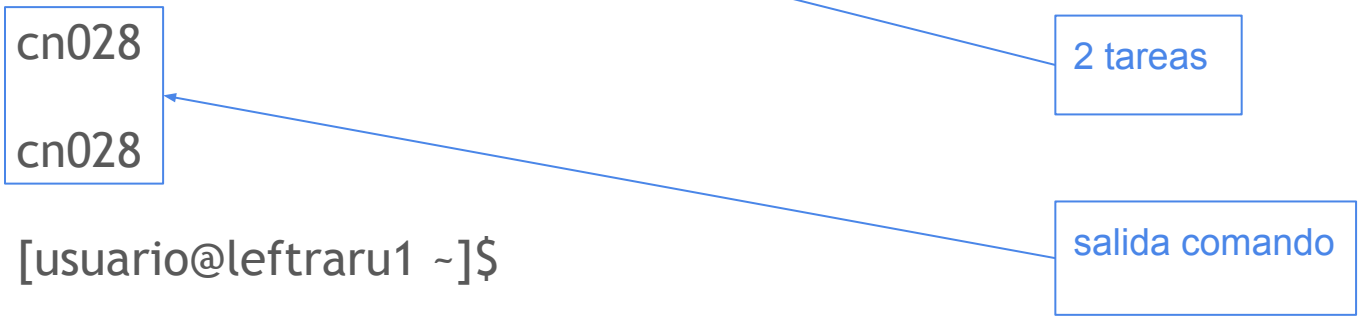
- Tres funciones principales:
  - Asignación de recursos (exclusivos y no exclusivos)
  - Framework para iniciar, ejecutar y monitorear trabajos
  - Gestiona tareas manejando una cola de recursos
- Integra una base de datos para reportes históricos
- Puede reservar diferentes recursos: CPU, socket, nodo o incluso por RAM
  - 6 nodos o 120 CPUs en el caso de Leftraru
- Permite a los administradores modificar tareas en ejecución y realizar reservas

NLHPC
National Laboratory for High Performance Computing Chile

Slurm: Primeros pasos

Pruebas desde la consola:

```
[usuario@leftraru1 ~]$ srun -n 2 hostname
cn028
cn028
[usuario@leftraru1 ~]$
```



Slurm: Monitorear mis tareas desde consola

Monitorear desde la consola:

[usuario@leftraru1 ~]$ squeue

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)
4400799	slims	example	usuario	R	0:00	1	cn042

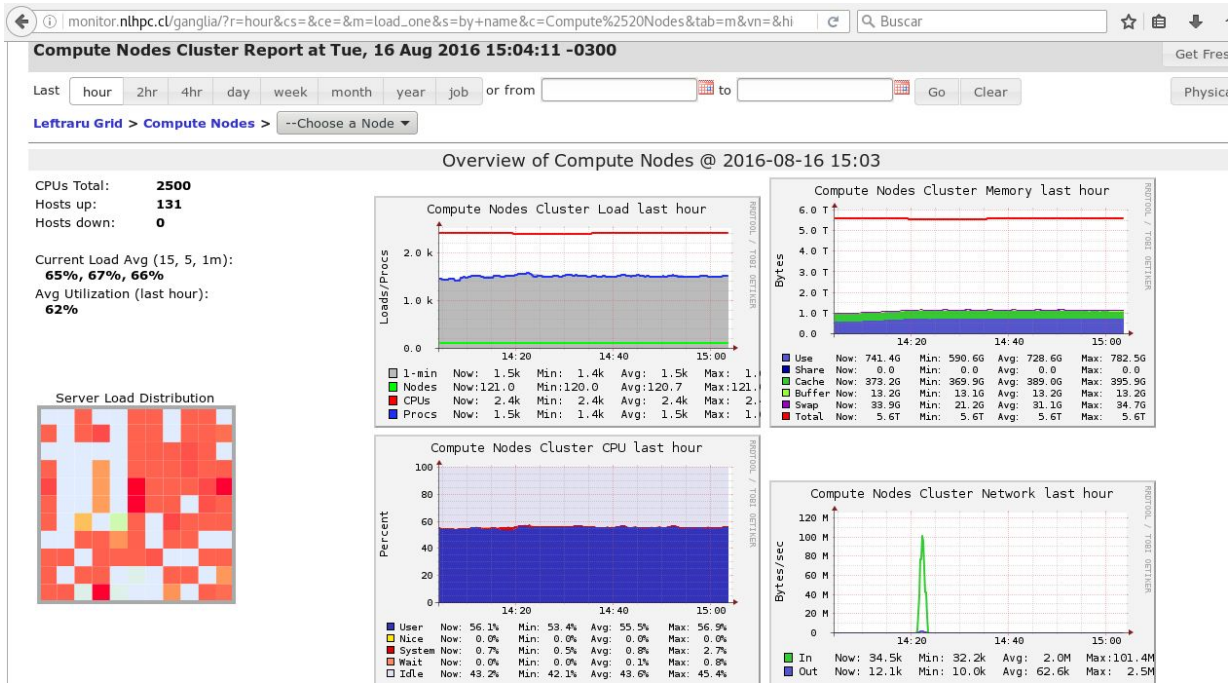
Obtener el nombre de el/los nodos y visitar:

http://monitor.nlhpc.cl/ganglia

slurm workload manager

NLHPC National Laboratory for High Performance Computing Chile

Slurm: Monitorear mis tareas desde la web



NLHPC
National Laboratory for High Performance Computing Chile

Slurm: Monitorear mis tareas desde la web

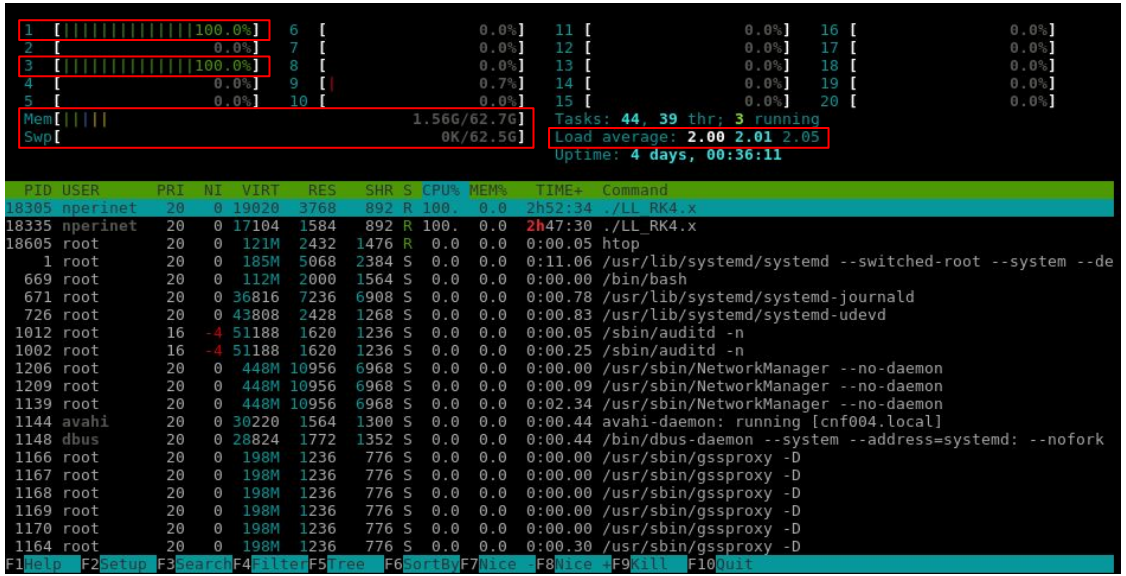
NLHPC
National Laboratory for High Performance Computing Chile

Slurm: Monitorear mis tareas desde la web

NLHPC
National Laboratory for High Performance Computing Chile

Monitorear mis tareas

Puede ingresar a través de ssh a un nodo en donde tenga una tarea en ejecución y ejecutar **htop**



NLHPC
National Laboratory for High Performance Computing Chile

Slurm: Scripts sbatch

Permiten:

- Ejecutar un scripts batch sin necesidad de estar siempre conectado
- Monitorear el estado de la tarea
- Monitorear recursos
- Monitorear estado de la cola
- Monitorear estado de las particiones

NLHPC
National Laboratory for High Performance Computing Chile

Módulos

Leftraru carga sus programas mediante “módulos”

- Permite varias versiones del mismo software
- No genera conflictos entre versiones
- El software está centralizado

Para cargar un módulo:

```
module load intel
```

module avail: muestra todos los nodos disponibles
module list: lista todos los módulos cargados
module unload: quita un módulo previamente cargado

NLHPC
National Laboratory for High Performance Computing Chile

SBATH Ejemplo de script básico test.sh

Utilizar su editor por consola preferido: vim, nano

```
#!/bin/bash
#SBATCH --job-name=ejemplo
#SBATCH --partition=slims
#SBATCH -n 1
#SBATCH --output=archivo_%j.out
#SBATCH --error=archivo_%j.err
#SBATCH --mail-user=usuario@gmail.com
#SBATCH --mail-type=ALL

module load intel

sleep 10
```

Ejecución:
sbatch test.sh

NLHPC
National Laboratory for High Performance Computing Chile

SBATCH Job Array

Slurm permite enviar y administrar millones de trabajos similares de una sola vez.

- Todos los trabajos deben tener las mismas condiciones iniciales
- Provee de variables para controlar la ejecución de los jobs

```
#!/bin/bash
#SBATCH --job-name=sleep-test
#SBATCH --partition=slims
#SBATCH -n 1
#SBATCH --output=st_%j.out
#SBATCH --error=st_%j.err
#SBATCH --array=1-10
#SBATCH --mail-user=usuario@dominio.cl
#SBATCH --mail-type=ALL

./ejecutable entrada_$SLURM_ARRAY_TASK_ID
```

NLHPC
National Laboratory for High Performance Computing Chile

SBATCH Ejemplo OpenMP

```
#!/bin/bash
#SBATCH --job-name=ejemplo
#SBATCH --partition=slims
#SBATCH --ntasks=1 # igual a parámetro “-n”
#SBATCH --cpus-per-task 20 # “-c”
#SBATCH --output=archivo_%j.out
#SBATCH --error=archivo_%j.err
#SBATCH --mail-user=usuario@gmail.com
#SBATCH --mail-type=ALL

export OMP_NUM_THREADS=20

./ejecutable
```

Ejecución:
sbatch test.sh

NLHPC
National Laboratory for High Performance Computing Chile

SBATCH Intel MPI

```
#!/bin/bash
#SBATCH --job-name=ejemplo
#SBATCH --partition=slims
#SBATCH -n 2
#SBATCH --output=ejemplo_impi_%j.out
#SBATCH --error=ejemplo_impi_%j.err
#SBATCH --mail-user=usuario@gmail.com
#SBATCH --mail-type=ALL

module load intel impi

srun ./hola_mundo
```

```
Hello from thread 00 out of 1 from process 00 out of 2 on
cn005
Hello from thread 00 out of 1 from process 01 out of 2 on
cn005
```

NLHPC
National Laboratory
for High Performance
Computing
Chile

Límites generales cuentas de usuario

- 120 CPUs
- 80 GB almacenamiento Lustre
- Walltime 3 días

Solicitud de cuentas:

visitar: http://www.nlhpc.cl (Servicios > Servicios para la Academia > Formulario)

ó solicitar información a info@nlhpc.cl

NLHPC
National Laboratory
for High Performance
Computing
Chile

SSH: Secure Shell - Introducción

- Protocolo seguro de acceso remoto a máquinas
- El cliente también se llama SSH
- Permite:
  - Ejecutar órdenes
  - Redirigir las “X” (sistema ventanas Unix, Linux)
  - Transferir archivos mediante FTP cifrado (sftp)
  - Transferencia de archivos bidireccional (scp)
  - Túneles de conexión, entre otras cosas
- Funciona comúnmente en el puerto TCP 22
- Admite múltiples tipos de cifrado

NLHPC
National Laboratory for High Performance Computing Chile

Funcionamiento Protocolo ssh

1. Se determina identidad de cliente y servidor
2. Establecimiento de canal seguro (cifrado 256 bits)
3. Cliente inicia sesión (autenticación) en el servidor

Dos métodos de autenticación:

- Por clave: mediante credenciales (usuario y password)
- Por llave: el cliente instala su llave pública en el servidor

NLHPC
National Laboratory for High Performance Computing Chile

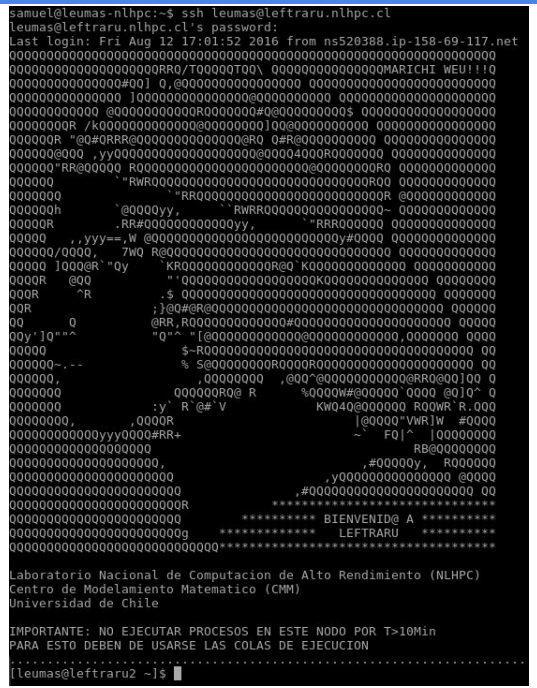
Instalación ssh

- Linux y Mac OSX: clientes integrados
- Windows:
  - Putty: cliente SSH: http://www.putty.org
  - WinSCP: transferencia de archivos: http://winscp.net

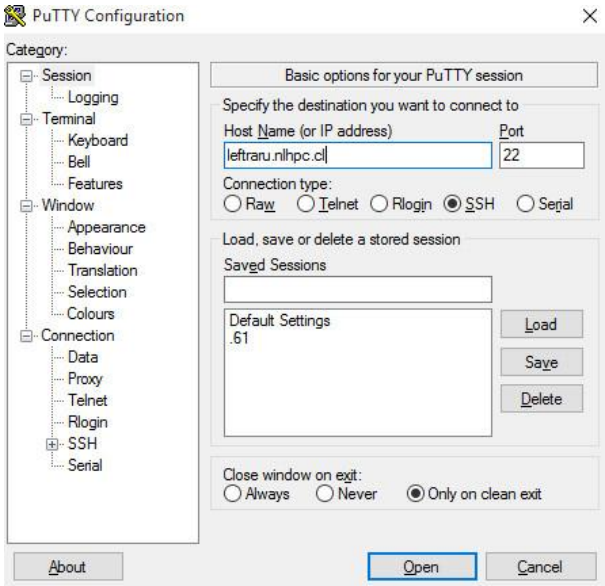
NLHPC
National Laboratory for High Performance Computing Chile

Iniciar Sesión ssh en leftraru

Linux, OSX:
ssh USER@leftraru.nlhpc.cl



Windows:



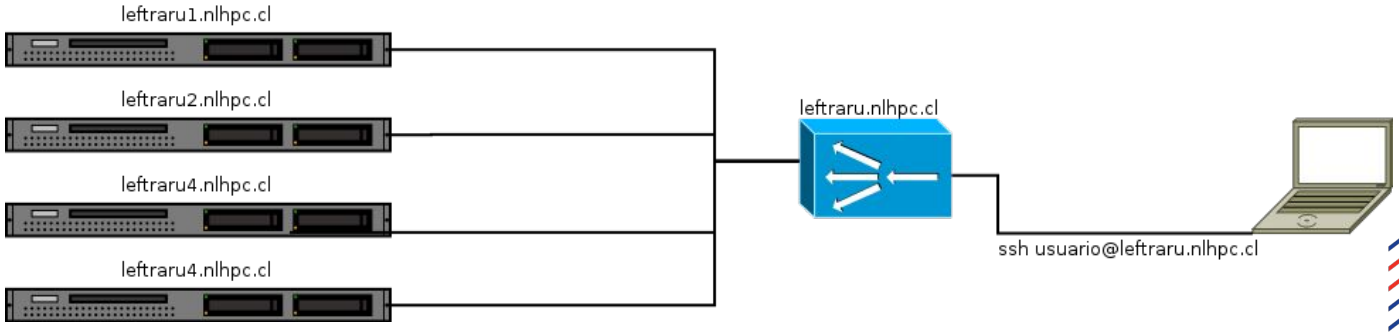
NLHPC
National Laboratory for High Performance Computing Chile

Consideraciones antes de conectar

- Balanceo de carga ssh:

  El usuario siempre hará login al nodo menos utilizado (OJO: uso de screen)

- Límite de trabajos en nodos login: 10 minutos
  - Esto afecta también a las transferencias de archivos grandes



NLHPC
National Laboratory for High Performance Computing Chile

ssh: copia de archivos desde línea de comandos

ssh permite copiar desde y hacia equipos remotos

scp archivo.txt usuario@leftraru.nlhpc.cl:~/ copia desde local hacia el servidor

scp usuario@leftraru.nlhpc.cl:~/archivo.txt ~/ copia desde el servidor hacia carpeta local

scp -C archivo.txt usuario@leftraru.nlhpc.cl:~/ "-C" utiliza compresión, envío más rápido

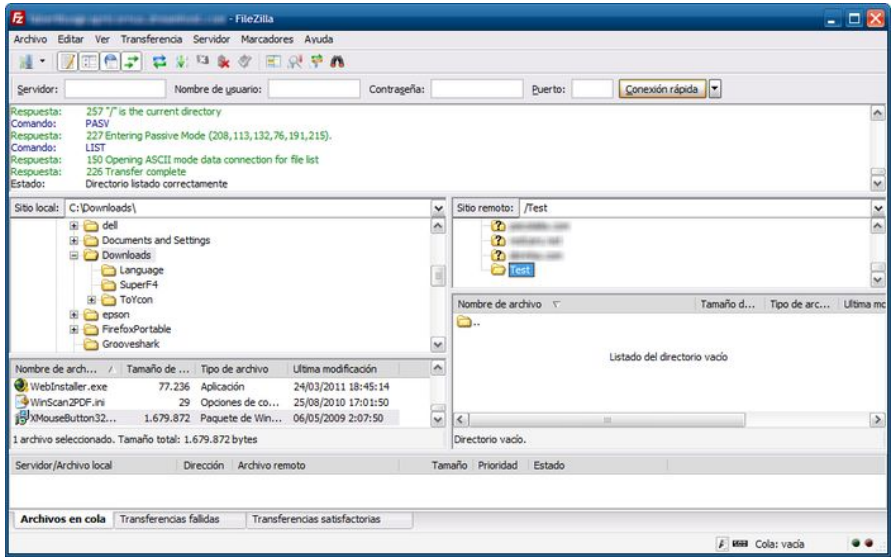
scp -r Directorio usuario@leftraru.nlhpc.cl:~/ copia un directorio completo

Nota: ~/ ruta relativa, indica la **raíz del home** del usuario

NLHPC
National Laboratory for High Performance Computing Chile

ssh copia de archivos GUI

Filezilla: http://filezilla-project.org



Otras opciones:

- WinSCP
- Midnight Commander (mc)

NLHPC
National Laboratory for High Performance Computing Chile

bash: screen

Cada vez que cerramos sesión, todos los procesos abiertos se cierran.
Incluso si los procesos fueron puestos en background.

Screen “desacopla” una sesión completa y la envía al fondo, lo que permite volver a retomarla incluso si la conexión se perdió.

Sólo debe ejecutar screen y luego ejecutar lo que se requiera.
Para desacoplar sólo debe presionar la siguiente secuencia: **control + a** y luego **d**

Para listar las sesiones abiertas: screen -list
Para remontar una sesión: screen -r **id**, donde id es el id de la sesión

Nota: leftraru cuenta con 4 nodos de login, por lo que usted debe recordar donde ejecutó screen

NLHPC
National Laboratory for High Performance Computing Chile

Fin Introducción: Manos a la obra

¿DUDAS?

NLHPC
National Laboratory
for High Performance
Computing
Chile

Ejercicios básicos:

Ejecute el siguiente comando: cp -r /home/courses/ejemplos/ .

1.- Ejecute:
- srun --reservation=curso -N 1 hostname
- srun --reservation=curso -N 2 hostname
- srun --reservation=curso -N 3 hostname

2.-
- Cree un script bash que ocupe 40 cores y ejecute el comando sleep 1000
- Lance el script
- Intente lanzarlo nuevamente ¿qué sucede?
- scancel: cancelar tareas

3.- Cree un script que reserve 1 nodo completo de forma exclusiva y ejecute el comando sleep 1000
4.- Cree un script que lance 4 trabajos, pero sólo dos por cada nodo
5.- Cree un script que lance un trabajo de un proceso en un nodo “fat”

```
Un amigo fiel: man <- sistema de ayuda en los programas linux
man sbatch
```

NLHPC
National Laboratory for High Performance Computing Chile

Ejercicio 6: job array

En el siguiente ejercicio usted jugará con la precisión de cálculo del número pi
Ingrese el directorio ejemplos/ejercicio_6, cree el siguiente script y luego ejecútelo:

```
#!/bin/bash
#SBATCH --job-name=pi-test
#SBATCH --partition=slims
#SBATCH -n 1
#SBATCH --output=st_%j.out
#SBATCH --error=st_%j.err
#SBATCH --array=1-10
#SBATCH --mail-user=usuario@mail.cl
#SBATCH --mail-type=ALL
#SBATCH --reservation=curso

module load intel impi
PRECISION=$( echo "$SLURM_ARRAY_TASK_ID*100000000" | bc )
srun ./pi_mpi.exe $PRECISION
```

Vigile la salida: watch -n 1 squeue
Una vez terminado los procesos, analice los archivos de salida (cat *.out)
¿Nota alguna diferencia?

NLHPC
National Laboratory for High Performance Computing Chile

Ejercicio 7: Intel - MPI

1.- Ingrese a la carpeta ejemplos/ejercicio_7

2.- Compile el programa hello.c con intel mpi
    mpiicc hello.c -o hello -fopenmp
    Ejecútelo con srun -n 1
    Ejecútelo con srun -n 20
    Ejecútelo con srun -n 21
    Ejecútelo con srun -c 1
    Ejecútelo con srun -c 2
    Ejecútelo con srun -c 20
    Ejecútelo con srun -c 21 (¿qué ocurre?)
    Ejecútelo con srun -n 1 -c 20
    Ejecútelo con srun -n 2 -c 20

Ejercicio 8: Intel - MPI

1.- Ingrese a la carpeta ejemplos/ejercicio_8

2.- Compile el programa hello.c con intel mpi

```
mpiicc hello.c -o hello -fopenmp
```

3.- Ejecute la aplicación con 2 procesos MPI y 20 OpenMP

4.- Analice los archivos de salida

NLHPC
National Laboratory for High Performance Computing Chile

Ejercicio 9: SBATCH Control tareas por RAM

- Ingrese al directorio ~/ejemplos/ejercicio_9
- Vea el contenido el ejemplo script.sh (**cat script.sh**)
- Ejecute y vigile la tarea
  - ¿Cuántos nodos ocupa?
  - ¿Cuanta RAM ocupa?
  - **Cancele la tarea**
- Edite nuevamente **script_slurm.sh**
  - Añada la línea: #SBATCH --mem-per-cpu=8192
- Vuelva a lanzar **script_slurm.sh**
  - ¿Qué ocurre?
  - ¿Cuanta RAM ocupa?
  - ¿Cuántos nodos ocupa ahora la tarea?

NLHPC
National Laboratory for High Performance Computing Chile

Ejercicio 10: Compilación - mkl

Intel MKL es una librería de optimización matemática que utiliza vectorización para mejorar el rendimiento

1.- Ingrese a la carpeta ejemplos/ejercicio_10 y liste los archivos (ls -l)

2.- Ingrese a matrix/linux

- El makefile de este ejemplo puede producir dos binarios: icc y gcc
- Para producir el binario con gcc: make gcc, producirá un archivo matrix.gcc
- Para producir el binario con icc: make gcc, producirá un archivo matrix.icc
- Ejecute ambos binarios
  - ./matrix.gcc
  - ./matrix.icc

3.- Retroceda a ingrese al directorio matrix/linux_mkl (cd ../linux_mkl)

- El makefile de este ejemplo puede generar un binario con intel MKL
- Para producir el binario con mkl: make mkl, producirá un archivo matrix.mkl
- OMP_NUM_THREADS es una variable que define la cantidad de procesos openmp
- export OMP_NUM_THREADS=1
- ./matrix.mkl

NLHPC
National Laboratory for High Performance Computing Chile

Ejercicio 10: Compilación - mkl

Ejercicio práctico:

1. Cree un script sbatch que ejecute el binario matrix.mlk, utilizando 1 proceso y 20 threads openmp
2. Modifique su script para lanzar ahora 2 procesos y 20 threads openmp

Hilos openmp: export OMP_NUM_THREADS=XX donde XX es el número de hilos

Compare los resultados

NLHPC
National Laboratory for High Performance Computing Chile

Ejercicio 11: Reservas mal hechas

- Ingrese a la carpeta ~/ejemplos/ejercicio_11
- Analice el archivo script.sh
  - ¿Cuántas CPU está reservando?
- Ejecútelo
- Vigile su tarea con ganglia
  - ¿Cuántas CPU está utilizando?
  - Cancele la tarea
- Modifique su script para que utilice las CPU que corresponden

NLHPC
National Laboratory for High Performance Computing Chile

Compilación y Optimización

- Ingrese a la carpeta “ejemplos” de su home
- Cargue el módulo intel
- Compile el ejemplo pi.c sin ningún flag de optimización y ejecútelo
- Compile el ejemplo pi.c con el flag “-O3” y ejecútelo
  - ¿Cuál fue la diferencia?
- Compile el ejemplo pi_openmpi.c
  - Ejecútelo con 1 sólo proceso
  - Ejecútelo con 10 procesos
  - Ejecútelo con 60 procesos
  - ¿Cuál es la diferencia?

NLHPC
National Laboratory for High Performance Computing Chile

FIN

¿DUDAS?

soporte@nlhpc.cl

NLHPC
National Laboratory
for High Performance
Computing
Chile